

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re Patent Application of	)	
	)	
Scott Shepard et al.	)	Group Art Unit: 2626
	)	
Application No.: 10/610,684	)	Examiner: Dorothy S. Siedler
	)	
Filed: July 2, 2003	)	
	)	
For: SYSTEMS AND METHODS FOR	)	
AIDING HUMAN TRANSLATION	)	

**APPEAL BRIEF**

Mail Stop: APPEAL BRIEF - PATENT  
Commissioner for Patents  
P.O. Box 1450  
Alexandria, Virginia 22313-1450

Sir:

This Appeal Brief is filed under Rule 41.37, appealing the final decision of the Examiner dated January 24, 2008. A Notice of Appeal was filed concurrently herewith. Each of the topics required by Rule 41.37 is presented herewith and labeled in accordance therewith. Only one copy of this Appeal Brief is required, in accordance with MPEP §1205.02.

**TABLE OF CONTENTS**

REAL PARTY IN INTEREST.....	PAGE 3
RELATED APPEALS AND INTERFERENCES.....	PAGE 4
STATUS OF CLAIMS.....	PAGE 5
STATUS OF AMENDMENTS.....	PAGE 6
SUMMARY OF CLAIMED SUBJECT MATTER.....	PAGES 7-11
GROUND OF REJECTION TO BE REVIEWED ON APPEAL.....	PAGE 12
ARGUMENTS.....	PAGES 13-26
SHIOTANI DISCLOSURE.....	PAGE 13
SCHULZ DISCLOSURE.....	PAGE 14
ARGUMENT #1 .....	PAGES 14-20
ARGUMENT #2 .....	PAGES 21-26
CONCLUSION.....	PAGE 27
CLAIMS APPENDIX.....	PAGES 28-40
EVIDENCE APPENDIX.....	PAGE 41
RELATED PROCEEDINGS APPENDIX.....	PAGE 42

**I. REAL PARTY IN INTEREST**

The real party in interest of the present application, solely for purposes of identifying and avoiding potential conflicts of interest by board members due to working in matters in which the member has a financial interest, is Verizon Communications Inc. and its subsidiary companies, which currently include Verizon Business Global, LLC (formerly MCI, LLC) and Cellco Partnership (doing business as Verizon Wireless, and which includes as a minority partner affiliates of Vodafone Group Plc). Verizon Communications Inc. or one of its subsidiary companies is an assignee of record of the present application.

**II. RELATED APPEALS AND INTERFERENCES**

There are no appeals or interferences related to the present application of which the Appellants are aware.

**III. STATUS OF CLAIMS**

Claims 1-47 are currently pending in the application and all stand finally rejected.

**IV: STATUS OF AMENDMENTS**

Subsequent to the final Office Action of January 24, 2008, (hereinafter “final Office Action”), Appellants have not filed an after-final Reply under 37 C.F.R. §1.116. The last amendment in this application was filed on November 15, 2007 responsive to a non-final office action. Accordingly, there are no outstanding amendments in this application.

#### **V. SUMMARY OF CLAIMED SUBJECT MATTER**

The following summary of the presently claimed subject matter indicates that certain portions of the specification (including the drawings) provide examples of embodiments of elements of the claimed subject matter. It is to be understood that other portions of the specification not cited herein may also provide examples of embodiments of elements of the claimed subject matter. It is also to be understood that the indicated examples are merely examples, and the scope of the claimed subject matter includes alternative embodiments and equivalents thereof. References herein to the specification are thus intended to be exemplary and not limiting.

In overview of the claimed subject matter, Appellants teach a human-language translation system in which a human-being translator does the translating from a first language (e.g., Spanish) to a second language (e.g., English). The human translator listens to an audio message in the first language while reading a text-transcript of that same audio message in that first language on a portion of a split screen display. The audio message and its corresponding text-transcript are to be translated by that person into the second language, in this example from Spanish to English. Each word in the text transcript of that first language is highlighted on the display screen in synchrony with the utterance of that word as it is spoken in the corresponding audio message. This serves as a translation-aid to the human translator as he, or she, types the translated message in the second language on another portion of the split screen. (See, e.g., Fig. 10) There is no machine-translating involved in Appellants' claimed subject matter - only machine-transcribing (transcribing from audio to text in the same language). With this overview

in mind, consider the claimed subject matter in detail. Appellants hereby map all independent claims to the drawings and Specification.

Independent claim 1 recites a method for facilitating translation of an audio signal that includes speech to another language, (e.g., at least Specification ¶'s [0008], [0009] and [0010] and Fig. 10) comprising:

retrieving a textual representation of the audio signal; (e.g., at least Specification ¶ [0044] and Fig. 4)

presenting the textual representation to a user; (e.g., at least Specification ¶'s [0045], [0046] and [0047] and Figs. 4-5)

receiving selection of a segment of the textual representation for translation; (e.g., at least Specification ¶ [0052] and Fig. 4)

obtaining a portion of the audio signal corresponding to the segment of the textual representation; (e.g., at least Specification ¶'s [0053] and [0059] and Figs. 4 and 8)

providing the segment of the textual representation and the portion of the audio signal to the user; (e.g., at least Specification ¶'s [0054] and [0065] and Figs. 4 and 8) and

receiving translation made by the user of the portion of the audio signal (e.g., at least Specification ¶ [0066] and Figs. 8-10).

Independent claim 20 recites a system for facilitating translation of speech between languages, (e.g., at least Specification ¶'s [0008], [0009] and [0010] and Fig. 10) comprising:



means for obtaining a textual representation of the speech in a first language (e.g., at least Specification ¶ [0044] and Fig. 4);

means for presenting the textual representation to a user (e.g., at least Specification ¶'s [0045], [0046] and [0047] and Figs. 4-5);

means for receiving selection of a portion of the textual representation for translation (e.g., at least Specification ¶ [0052] and Fig. 4);

means for retrieving an audio signal in the first language that corresponds to the portion of the textual representation (e.g., at least Specification ¶'s [0053] and [0059] and Figs. 4 and 8);

means for providing the portion of the textual representation and the audio signal to the user (e.g., at least Specification ¶'s [0054] and [0065] and Figs. 4 and 8); and

means for receiving translation made by the user of the audio signal into a second language (e.g., at least Specification ¶ [0066] and Figs. 8-10).

Independent claim 21 recites a translation system (e.g., at least Specification ¶'s [0008] - [0010] and [0024] - [0025] and Figs. 1-3 and 10), comprising:

a memory configured to store instructions (e.g., at least Specification ¶ [0028] and Fig. 2); and

a processor configured to execute the instructions in memory (e.g., at least Specification ¶ [0028] and Fig. 2) to:

obtain a transcription of an audio signal that includes speech (e.g., at least Specification ¶ [0044] and Fig. 4),

present the transcription to a user (e.g., at least Specification ¶'s [0045], [0046] and [0047] and Figs. 4-5),

receive selection of a portion of the transcription for translation (e.g., at least Specification ¶ [0052] and Fig. 4),

retrieve a portion of the audio signal corresponding to the portion of the transcription (e.g., at least Specification ¶'s [0053] and [0059] and Figs. 4 and 8),

provide the portion of the transcription and the portion of the audio signal to the user (e.g., at least Specification ¶'s [0054] and [0065] and Figs. 4 and 8), and

receive from the user a translation made by the user of the portion of the audio signal (e.g., at least Specification ¶ [0066] and Figs. 8-10).

Independent claim 40 recites a graphical user interface (e.g., at least Specification ¶'s [0061], [0062] and [0066] and Figs. 9-10), comprising:

a transcription section that includes a transcription of non-text information in a first language (e.g., at least Specification ¶ [0062] and Figs. 9-10);

a translation section that receives a translation made by the user of the non-text information into a second language (e.g., at least Specification ¶ [0062] and Figs. 9-10); and

a play button (e.g., at least Specification ¶ [0062] and Figs. 9-10) that, when selected, causes:

retrieval of the non-text information to be initiated (e.g., at least Specification ¶'s [0063] - [0064] and Figs. 8-10),

playing of the non-text information (e.g., at least Specification ¶ [0064] and Figs. 8-10), and

the playing of the non-text information to be visually synchronized with the transcription in the transcription section (e.g., at least Specification ¶ [0065] and Figs. 8-10).

Independent claim 47 recites a method (e.g., at least Figs. 4 and 8), comprising:

a user listening to an audio playback of information in a first language while viewing a textual transcription of said information in said first language on a transcription section of a graphical user interface (GUI), said textual transcription being synchronized with said audio playback (e.g., at least Specification ¶'s [0065] - [0066] and Fig. 10); and

said user translating said audio playback of said information thereby obtaining a translation in a second language, said user using a different section of said GUI to display said translation while making said translation (e.g., at least Specification ¶'s [0065] - [0066] and Fig. 10),

whereby the synchronizing of said audio playback with said textual transcription aids said user in making said translation (e.g., at least Specification ¶'s [0070] - [0071]).

## **VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL**

In the final Office Action, the following rejections were made:

### **Ground Number One:**

Claims 1-40, 42-45 and 47 are rejected under 35 U.S.C. § 103(a) as being un-patentable over Shiotani (U.S. 4,814,988) in view of Schulz (U.S. 6,360,237).

### **Ground Number Two:**

Claims 41 and 46 are rejected under 35 U.S.C. § 103(a) as being un-patentable over Shiotani in view of Schulz as applied to claim 40 and further in view of Saindon (U.S. 6,820,055).

These are the sole grounds of rejection in the final Office Action, but only Ground Number One is to be reviewed on appeal.<sup>1</sup>

---

<sup>1</sup> As noted in the Argument section which follows, all independent claims shall stand or fall with claim 1. Furthermore, Appellants' dependent claims shall stand or fall with their independent claims. Therefore, the second ground of rejection is moot.

## **VII: ARGUMENT**

The independent claims on appeal are claims 1, 20, 21, 40 and 47. Appellants shall let their dependent claims stand or fall with their respective independent claims and let independent claims 20, 21, 40 and 47 stand or fall with claim 1. Therefore, the only claim for which arguments are being presented below is claim 1. All grounds of rejection are moot but for ground #1.

### **SUMMARY OF THE SHIOTANI DISCLOSURE:**

Shiotani relates to a machine translation system for translating all or a selected portion of an input sentence. (title) The Shiotani input is derived from non-audio sources such as an optical character reader (OCR). (col. 2, lines 11-14) Shiotani, therefore, does not teach translation of language presented in an audio format, as conceded in the Office Action (see Office Action, pg. 4, bottom)

Shiotani presents a block diagram of essential parts of its machine translation system in its Fig. 1. (col. 1, lines 66-67). Translating part 7, shown in Shiotani's Fig. 1, is the computer mechanism that does the actual translating; it translates the content in original buffer 6 by operation of a dictionary look-up/morpheme analyzing function, a syntax analyzing function, a transforming function and a generating function. (col. 2, lines 30-34) A translation buffer 10 is provided for storing the result of the translation. (col. 2, lines 38-39) In addition, a correcting means 11 is used by the human operator to correct the translation result that is displayed on a terminal screen. (col. 2, lines 39-41) Accordingly, Shiotani teaches machine translation of non-

audio human language and human correction of that translation.

**SUMMARY OF THE SCHULZ DISCLOSURE:**

Schulz relates to a method and system for performing text edits during audio recording playback for transcription. (title and col. 1, lines 7-9) Schulz does not disclose, or relate to, translation from one language to a different language but is limited to transcribing from audio to text within the same language. Schulz discloses a method for editing (correcting) written text in a particular language in a text editor which automatically aligns a cursor in the written text on a screen with a particular spoken word in that same language during playback of an audio recording. (col. 2, lines 48-51)

**I. CLAIM 1 IS ALLOWABLE BECAUSE SHIOTANI AND/OR SCHULZ DO NOT DISCLOSE OR SUGGEST ALL CLAIM ELEMENTS**

With respect to claim 1, rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over Shiotani in view of Schulz, Appellants' first argument in this appeal is that all claim elements of claim 1 are not disclosed or suggested by Shiotani or Schulz taken individually or in any reasonable combination.<sup>2</sup> Claim 1 recites a method that facilitates translation of an audio signal that includes speech to another language comprising *inter alia* the act of "receiving translation made by the user of the portion of the audio signal." Clearly, Appellants' claim 1 is limited, *inter alia*, to translation made by a user who is a human being

---

<sup>2</sup> Appellants further contest the Office Action's combining of these references in the first place, subsequently discussed hereinbelow.

translator who listens to the audio and translates it from the spoken human language that he/she hears into a different human language. This translation operation is directly discernable from this claim element and throughout Appellants' specification, including drawings and other claims.

Against this claim limitation directed to translation, the Office Action associates one, and only one, cite from only the Shiotani reference, namely: "(column 2 lines 39-41, *the user provides correction of the translation result of the specified input region*)."

 (Office action, pg 4; emphasis in original) The Office Action, therefore, does not allege that the other cited reference, Schulz, teaches or suggests this claim limitation, and Appellants agree. However, Appellants respectfully disagree that Shiotani teaches or suggests this claim limitation because any translation that is accomplished in Shiotani is performed by operation of a machine.

After all, Shiotani is entitled: "Machine Translation System Translating All Or A Selected Portion Of An Input Sentence." In Fig. 1 of Shiotani, the translation is performed in "translating part" 7. Shiotani states in column 2, lines 30-41 (emphasis added):

A translating part 7 for translating the content in said original buffer 6, is composed, for example, of a dictionary look-up/morpheme analyzing part, a syntax analyzing part, a transforming part, and a generating part. Numeral 8 is a changeover part to change over the grammatical rules on a grammatical rule table 9 applied to the translating operation of said translating part 7 depending on the state of flag 5. A translation buffer 10 for storing the result of translation, and a correcting means 11 used by the operator to correct the translation result displayed on the CRT are provided.

From this section of Shiotani, it is clear that translation is accomplished by machine while correction of that translation result is accomplished by a human operator. The human operator in Shiotani, who interacts with the Shiotani translation system, is not a translator, but is a reviewer

who makes corrections in the same language being reviewed.<sup>3</sup>

Appellants respectfully submit that reviewing text in a resultant, translated language and making corrections to that text in that same language cannot be viewed as translation by any reasonable interpretation of that term. The relevant dictionary definition of translation is: "an act, process, or instance of translating: as a: a rendering from one language into another; also: the product of such a rendering" (Merriam Webster's Collegiate Dictionary, Tenth Edition). The dictionary definition of translation requires change from one language to a different language. Clearly, translation in the context of the instant patent application also requires going from one human language to a different human language. Shiotani teaches translation by way of machine and it simply does not teach or suggest that a human being is translating from one language to another. Shiotani's only disclosed human operator is characterized as only correcting the result of the translation and, therefore, operates upon and within the resultant language. To conclude otherwise is to impermissibly read more into Shiotani that it is disclosing.

The Office Action, therefore, is impermissibly reading more into Shiotani than it is disclosing when it states: "However the examiner respectfully disagrees, and contends that the correction step taken by the user is in fact part of the translation process. The user observes the source utterance, then the target utterance (Figure 4(a) and 4(b)) and determines that the target

---

<sup>3</sup> In another reference to "operator" in column 1, lines 7-13 of Shiotani, it discusses an "interactive" method between operator and system, without specifying what that interaction is. And, aside from its claims, the only other reference to "operator" in Shiotani, is in column 1, lines 40-41, where it mentions enhancing the "controllability" of the operator but doesn't specify what that means. The only possible explanation of what these terms mean is obtainable from column 2, line 40, i.e., related to "correction" of the translated result.



utterance is incorrect, i.e. incorrectly translated, and provides the correct translation.” (Office Action, pg 2)

In rebuttal to the first sentence of this Office Action statement, Appellants point out that Shiotani's correction step is not part of the translation process, per se, because the correction step is performed through correction means 11 which contributes to system operation only after completion of the translation process by operation of translating part 7 (*see* Shiotani Fig. 1). The translation had been previously accomplished in part 7 and correction is made by the human operator to the resultant, translated language by way of correcting means 11 in the same resultant, translated language. This amounts to editing, not translating.

In rebuttal to the second sentence of this Office Action statement, the user may view the source and target utterances, but the language “to correct the translated result” as specified in Shiotani does not mean “to translate” in any event. In order to read Shiotani on Appellants' claim 1, one first needs to unreasonably redefine “correct the translated result” as being equivalent to “translate.” Otherwise, there is not even a colorable operator-translation activity in Shiotani because it is a machine translation system.

*Arguendo*, if one were to choose to ignore dictionary definitions, common usage, logic and reason and thereby assume that “correction” means “translation” which the Office Action appears to be asserting and with which Appellants disagree, Shiotani would still not read on Appellants' claim 1 anyway. Appellants' “receiving translation” claim element says: “receiving translation made by the user of the portion...” Appellants' claimed user makes a translation of

the "portion" and this, of course, can only mean a translation of the entire portion. Appellants' human translator cannot limit its translation effort to translating only a fraction of the recited portion because there is no other translating mechanism provided in Appellants' disclosure. Therefore, the human translator must translate the entire recited portion in order to achieve a "translation made by the user of the portion" as recited in claim 1.

But, the correcting (*arguendo* translating) that is being made by the human operator in Shiotani is not a correction of an equivalent to Appellants' entire portion. Rather, it is a correction of only incorrect aspects, if any, of the Shiotani sentence, or portion thereof, because one cannot correct what is already correct. Therefore, even when erroneously equating translating with correcting to give Shiotani maximum advantage in its role as a cited reference, if any part of a Shiotani sentence, or sentence portion, that is subject to Shiotani operator correction is a correct part, then the Shiotani operator does not correct (*arguendo*, translate) the correct part wherefore Shiotani does not correct (*arguendo*, translate) the entire sentence, or sentence portion, and does not read on this element of Appellants' claim 1.

The only unreasonable way that Shiotani could arguably read on this claim element would be to first irrationally postulate that *correction means translation* and, thereafter, building on this arbitrary redefinition, interpret that Shiotani necessarily teaches that the entire part of a Shiotani sentence, or portion thereof, that is subject to correction needs translation correction. But, if Shiotani actually taught this, then Shiotani would be teaching that its machine translation system is an utter failure because its machine translation system would then never be translating anything correctly! Therefore, under these hypothetical conditions, a patent should not have

been issued in the first place because the Shiotani application would then not have disclosed an operative embodiment! Obviously, Shiotani is not teaching that its machine translation system is an utter failure which is strong evidence that the Office Action's interpretation of Shiotani as applied to claim 1 is erroneous.

As noted above, this *arguendo* argument is premised on an impermissible re-defining of correction as being equivalent to translation. This is plainly inaccurate on its face. Shiotani does not disclose a limitation on its corrections to a particular class of corrections in the resultant language. Therefore, the Shiotani operator can correct for any error such as, e.g., spelling errors, punctuation errors, formatting errors, etc. in the resultant language which corrections clearly have nothing to do with the subject of translation. Making corrections for these kinds of errors in the translation result are clearly not, in and of itself, performing translations. Thus, correction is not translation and correcting a translation result need not be a translation.

Therefore, the only conclusion that one can draw from the above analysis of the Office Action's stated position is that the Office Action's interpretation of Shiotani's translation correction function as allegedly reading on Appellants' "receiving translation made by the user of the portion of the audio signal" step of claim 1 is incorrect. To correct the translation result is not translation.

Finally, for sake of completeness, regardless of the inability of the references to be combined (which shall be discussed below), consider if Schulz could cure this deficiency in Shiotani. Schulz does not teach or suggest an operation involving a human translator but, rather, teaches a human transcriber. The term "translation" or "translating" does not appear in Schulz

at all. However, "transcriptionist" does appear frequently in Schulz; see, for example, column 1, lines 23, 41, 47, 54, 62 and 66; column 2, lines 10, 18, 29 and 41. Clearly Schulz is directed to an activity in which its disclosed human operator is merely transcribing from language A to language A which is not translating. Therefore, hypothetically combining Schulz with Shiotani, for argument's sake, also does not produce a combined disclosure that teaches or suggests a human translator operation. Moreover, as noted, the Office Action does not even attempt to apply Schulz against this claim element in the first place. Therefore, Schulz and Shiotani, taken individually or in any reasonable combination, do not disclose or suggest "receiving translation made by the user of the portion of the audio signal" as recited in claim 1.

Claim 20 recites, *inter alia*, "means for receiving translation made by the user of the audio signal into a second language" and this is not disclosed or suggested by Shiotani and/or Schulz for reasons similar to those given above for claim 1.

Claim 21 recites, *inter alia*, "a processor configured to execute the instructions in memory to...receive from the user a translation made by the user of the portion of the audio signal" and this is not disclosed or suggested by Shiotani and/or Schulz for reasons similar to those given above for claim 1.

Claim 40 recites, *inter alia*, "a translation section that receives a translation made by the user of the non-text information into a second language" and this is not disclosed or suggested by Shiotani and/or Schulz for reasons similar to those given above for claim 1.

Claim 47 recites, *inter alia*, "said user translating said audio playback of said information thereby obtaining a translation in a second language, said user using a different section of said

GUI to display said translation while making said translation” and this is not disclosed or suggested by Shiotani and/or Schulz for reasons similar to those given above for claim 1.

## **II. SHIOTANI AND SCHULZ ARE NOT PROPERLY COMBINABLE**

The Office Action concedes that Shiotani does not disclose or suggest the audio signal recited in claim 1. (Office Action, page 3) Appellants agree.

The Office Action then presents Schulz which discloses audio transcription but has absolutely nothing to do with translation and immediately concludes that, because Schulz (1) mentions in its background section that automatic speech recognition systems convert spoken language to written text and (2) discloses the synchronizing of text with a specific spoken word during playback of an audio file, it would be obvious to one of ordinary skill in the art at the time of the invention to combine Schulz with Shiotani to read on Appellants’ subject matter as recited in claim 1. The alleged rationale given is: “it would have been obvious to one of ordinary skill in the art at the time of the invention to retrieve a textual representation of an audio signal for translation in Shiotani, since it would enable the system to translate spoken language as well as textual documents.” (Office Action, pg 3) Appellants respectfully disagree that this is satisfactory rationale at least for the reason that this is no more than a conclusory statement that merely recites advantages offered by Appellants’ claimed subject matter, those advantages being apparent in hindsight after one reads Appellants’ claims.

The Office Action continues that it would also have been obvious to provide a segment of text and a corresponding portion of audio to the user in Shiotani because Schulz’s text editor can

edit text for transcription or translation purposes, and that “the combination of [Schulz’s] text editing software with [Shiotani’s] standard machine translation system would produce the predictable result of enabling the user to quickly and easily edit, or translate, text displayed on the monitor without interruption during playback of the speech from an audio recording” (Office Action, pgs. 3-4) Appellants again respectfully disagree that this is satisfactory rationale for finding obviousness at least for the reason that this is also no more than a conclusory statement that is also merely reciting advantages offered by Appellants’ claimed subject matter, those advantages being apparent in hindsight after reading Appellants’ claims.

Appellants rely on the recently decided case KSR International Co. v. Teleflex Inc., 550 U.S. \_\_\_\_\_ (April 30, 2007) (citing In re Kahn, 441 F.3d 977, 988 (Fed. Cir. 2006)), (hereinafter “KSR”) where it was held that rejections on obviousness grounds cannot be sustained by mere conclusory statements; instead, there must be some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness. Appellants submit that the above-noted statements in the Office Action do not represent articulated reasoning. The Examiner’s purported motivation to combine the cited references is merely conclusory and based on impermissible hindsight. If it were as obvious to have combined the teachings of Shiotani and Schulz to achieve the alleged “predictable result” as the Office Action represents, Appellants query, as a threshold matter, why that combination has not previously been made. The answer to this query is that the combination is actually not obvious, at least because there are multiple differences between the two references including un-related technological disciplines, namely, optical character recognition versus audio technology, and that only after reading Appellants’

claims may the combination arguably appear to be obvious. After all, the Examiner has conducted a thorough search and, by not finding a description of that combination within a single reference, has shown that the alleged “predictable result” has apparently not yet been produced in tangible form.

In this connection, MPEP 2141 (III) offers guidance with respect to various rationales to support rejections under KSR. One exemplary rationale is “obvious to try - choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success.” Appellants submit that it is not obvious to try to combine Shiotani and Schulz for several reasons. First of all, Shiotani is a machine language-translation system for operating exclusively on text, involving a human operator only for correction purposes; this reference does not even hint at audio data input. Quite differently, Schulz is a transcribing system for editing exclusively a transcription of audio (voice) with synchronization between the spoken language and the transcription; this reference does not even hint at language-translation or textual data input. Appellants submit that translation between two different languages on the one hand and transcription from one media to another in the same language on the other hand are two very different activities and common sense suggests that there is no motivation to be derived from a reading of either of these references to seek its combination with the other.

In addition, they operate with divergent technologies, where their combination offers no predictable solution and no reasonable expectation of success. There are divergent technologies involved in, and resultant divergent skill requirements needed for handling, (1) conversion of Shiotani's text via optics to digital signals for further processing, versus (2) conversion of

Schulz's audio signals to digital signals for further processing.

For example, Shiotani (col. 2, lines 13-14) discusses an optical character reader (OCR) involving principles based on the physics of optics. Momentarily expanding on this subject for illustrative purposes, OCR is mechanical or electronic translation of images of text into machine-editable text, using optical techniques such as mirrors and lenses in combination with scanners and digital processing. OCR is a process by which glyph images (the visual image of a character) yield character codes. Given a picture of letters arranged as words, OCR is supposed to give back strings of character codes arranged as words. Individual dots of the digital image are represented by a number that varies as function of black through gray to white (for black/white images). Locations of the scan are identified as pixels (picture elements). This brief snippet of OCR information may provide an inkling of what someone with skill in this art has mastered.

By contrast, Schulz (col. 4, lines 50-53) discusses a mu-law encoded eight-bit digital signal. The mu-law algorithm is a companding algorithm, whose purpose is to reduce the dynamic range of an audio signal. In the analog domain, this can increase the signal-to-noise ratio achieved during transmission, and in the digital domain it can reduce quantization error. Beyond this, speech recognition involves many considerations such as complexity of the language model. By this is meant the number of permissible words following each word. The simplest language model can be specified as a finite-state network. One measure of the difficulty of the task of combining vocabulary size and language model is called "perplexity" which is the geometric mean of the number of words that can follow a word, after a language



model has been applied. This does not begin to scratch the surface of the subject of speech recognition, but this brief snippet of speech recognition information may provide an inkling of what someone with skill in this art has mastered.

Appellants have juxtaposed the above two paragraphs to clearly show that the subjects discussed therein are mutually exclusive. One topic has virtually nothing to do with the other. Accordingly, one skilled in the audio signal processing art need not be similarly skilled in the text signal processing and vice-versa. This clear difference in technologies, in addition to the translation/transcription difference noted above make it unlikely, in Appellants' view, to find any motivation within either of these references to combine one with the other.

The initial burden of establishing a *prima facie* basis to deny patentability to a claimed invention always rests upon the Examiner. In re Oetiker, 977 F.2d 1443, 24 U.S.P.Q.2d 1443 (Fed. Cir. 1992). In rejecting a claim under 35 U.S.C. § 103, the Examiner must provide a factual basis to support the conclusion of obviousness. In re Warner, 379 F.2d 1011, 154 U.S.P.Q. 173 (C.C.P.A. 1967). Based upon the objective evidence of record, the Examiner is required to make the factual inquiries mandated by Graham v. John Deere Co., 86 S.Ct. 684, 383 U.S. 1, 148 U.S.P.Q. 459 (1966). KSR International Co. v. Teleflex Inc., 550 U.S. \_\_\_\_\_ (April 30, 2007). The Examiner is also required to explain how and why one having ordinary skill in the art would have been realistically motivated to modify an applied reference and/or combine applied references to arrive at the claimed invention. Uniroyal, Inc. v. Rudkin-Wiley Corp., 837 F.2d 1044, 5 U.S.P.Q.2d 1434 (Fed. Cir. 1988). In view of the differences between the references that have been presented herein, Appellants respectfully submit that the Examiner has

not met these standards; for example, in this instance, the Office Action has not presented sufficient explanation of how and why one having ordinary skill in the art would have been realistically motivated to modify either applied reference and/or combine these applied references to arrive at the claimed subject matter. The Office Action merely presents advantages which become appreciated after a reading of Appellants' claims.

It is established law that one "cannot use hindsight reconstruction to pick and choose among isolated disclosures in the prior art to deprecate the claimed invention." *Ecolchem, Inc. v. Southern Cal. Edison Co.*, 227 F.3d 1361, 1371, 56 USPQ2d 1065 (Fed. Cir. 2000) (citing *In re Fine*, 837 F.2d 1071, 1075, 5 USPQ2d 1780, 1783 (Fed. Cir. 1988)). Indeed, "[c]ombining prior art references without evidence of such a suggestion, teaching, or motivation simply takes the inventor's disclosure as a blueprint for piecing together the prior art to defeat patentability ... the essence of hindsight." *In re Dembiczak*, 175 F.3d 994, 999, 50 USPQ2d 1614, 1617 (Fed. Cir. 1999). Appellants submit that in this instance Appellants' claim 1 was used as such a blueprint to piece together Shiotani and Schulz.

Schulz was cited to cure the deficiency of no audio disclosure within Shiotani, whereby that deficiency is not cured because the references cannot be combined for the reasons given above. For the foregoing reasons the 35 U.S.C. § 103(a) rejection of claim 1 should be REVERSED and the claim allowed.

The other independent claims, claims 20, 21, 40 and 47 should each be allowed for reasons that are the same as, or similar to, those given above with respect to claim 1.

All dependent claims are allowable, at least for reasons based on their dependencies from

allowable base claims.

**CONCLUSION**

For either one, or both, of the two distinct arguments presented above, Appellants respectfully request that the Honorable Board reverse the final rejection of the appealed claims.

To the extent necessary, a petition for an extension of time under 37 C.F.R. § 1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of this paper, including extension of time fees, to Deposit Account No. 07-2347 and please credit any excess fees to such deposit account.

Respectfully submitted,

/Eden Stright/ Eden Stright, Reg. No. 51,205, for  
Joel Wall - Registration 25,648

Date: March 31, 2008

Verizon  
Patent Management Group  
1515 Courthouse Road, Suite 500  
Arlington, VA 22201 - 2909  
Tel: 703.351.3586  
Fax: 703.351.3665  
**Customer No. 25537**

**VIII: CLAIMS APPENDIX**

1. A method for facilitating translation of an audio signal that includes speech to another language, comprising:
  - retrieving a textual representation of the audio signal;
  - presenting the textual representation to a user;
  - receiving selection of a segment of the textual representation for translation;
  - obtaining a portion of the audio signal corresponding to the segment of the textual representation;
  - providing the segment of the textual representation and the portion of the audio signal to the user; and
  - receiving translation made by the user of the portion of the audio signal.
2. The method of claim 1, wherein the retrieving a textual representation includes:
  - generating a request for information,
  - sending the request to a server, and
  - obtaining, from the server, at least the textual representation of the audio signal.
3. The method of claim 1, wherein the presenting the textual representation to a user, includes:

obtaining the audio signal,  
providing the audio signal and the textual representation of the audio signal to the user,  
and  
visually synchronizing the providing of the audio signal with the textual representation of  
the audio signal.

4. The method of claim 3, wherein the obtaining the audio signal includes:  
accessing a database of original media to retrieve the audio signal.

5. The method of claim 3, wherein the obtaining the audio signal includes:  
receiving input, from the user, regarding a desire for the audio signal,  
initiating a media player, and  
using the media player to obtain the audio signal.

6. The method of claim 1, wherein the receiving selection of a segment of the textual  
representation includes:  
identifying a portion of the textual representation selected by the user,  
accessing a server to obtain text corresponding to the portion of the textual  
representation, and  
receiving, from the server, the text corresponding to the portion of the textual  
representation.

7. The method of claim 6, wherein the text includes a transcription of the audio signal and metadata corresponding to the portion of the textual representation.

8. The method of claim 1, wherein the obtaining a portion of the audio signal includes:

initiating a media player, and

using the media player to obtain the portion of the audio signal.

9. The method of claim 8, wherein the using the media player includes:  
identifying, by the media player, the segment of the textual representation, and  
retrieving the portion of the audio signal corresponding to the segment of the textual representation.

10. The method of claim 9, wherein the identifying the segment includes:  
identifying time codes associated with a beginning and an ending of the segment of the textual representation.

11. The method of claim 9, wherein the segment of the textual representation includes a starting position in the textual representation; and  
wherein the identifying the segment includes:

identifying a time code associated with the starting position in the textual representation.

12. The method of claim 1, wherein the providing the segment of the textual representation and the portion of the audio signal to the user includes:

displaying the segment of the textual representation in a same window as will be used by the user to provide the translation of the portion of the audio signal.

13. The method of claim 1, wherein the providing the segment of the textual representation and the portion of the audio signal to the user includes:

visually synchronizing the providing of the portion of the audio signal with the segment of the textual representation.

14. The method of claim 13, wherein the segment of the textual representation includes time codes corresponding to when words in the textual representation were spoken.

15. The method of claim 14, wherein the visually synchronizing the providing of the portion of the audio signal with the segment of the textual representation includes:

comparing times corresponding to the providing of the portion of the audio signal to the time codes from the segment of the textual representation, and

visually distinguishing words in the segment of the textual representation when the words are spoken during the providing of the portion of the audio signal.



16. The method of claim 1, wherein the providing the segment of the textual representation and the portion of the audio signal to the user includes:

permitting the user to control the providing of the portion of the audio signal.

17. The method of claim 16, wherein the permitting the user to control the providing of the portion of the audio signal includes:

allowing the user to at least one of fast forward, speed up, slow down, and back up the providing of the portion of the audio signal using foot pedals.

18. The method of claim 16, wherein the permitting the user to control the providing of the portion of the audio signal includes:

permitting the user to rewind the portion of the audio signal at least one of a predetermined amount of time and a predetermined number of words.

19. The method of claim 1, further comprising:

publishing the translation to a user-determined location.

20. A system for facilitating translation of speech between languages, comprising:

means for obtaining a textual representation of the speech in a first language;

means for presenting the textual representation to a user;

means for receiving selection of a portion of the textual representation for translation;  
means for retrieving an audio signal in the first language that corresponds to the portion of the textual representation;  
means for providing the portion of the textual representation and the audio signal to the user; and  
means for receiving translation made by the user of the audio signal into a second language.

21. A translation system, comprising:  
a memory configured to store instructions; and  
a processor configured to execute the instructions in memory to:  
obtain a transcription of an audio signal that includes speech,  
present the transcription to a user,  
receive selection of a portion of the transcription for translation,  
retrieve a portion of the audio signal corresponding to the portion of the transcription,  
provide the portion of the transcription and the portion of the audio signal to the user, and  
receive from the user a translation made by the user of the portion of the audio signal.

22. The system of claim 21, wherein when obtaining a transcription, the processor is configured to:

- generate a request for information,
- send the request to a server, and
- obtain, from the server, at least the transcription of the audio signal.

23. The system of claim 21, wherein when presenting the transcription to a user, the processor is configured to:

- obtain the audio signal,
- provide the audio signal and the transcription of the audio signal to the user, and
- visually synchronize the providing of the audio signal with the transcription of the audio signal.

24. The system of claim 23, wherein when obtaining the audio signal, the processor is configured to:

- access a database of original media to retrieve the audio signal.

25. The system of claim 23, wherein when obtaining the audio signal, the processor is configured to:

- receive input, from the user, regarding a desire for the audio signal,
- initiate a media player, and

use the media player to obtain the audio signal.

26. The system of claim 21, wherein when receiving selection of a portion of the transcription, the processor is configured to:

identify a range of the transcription selected by the user,  
access a server to obtain text corresponding to the range of the transcription, and  
receive, from the server, the text corresponding to the range of the transcription.

27. The system of claim 26, wherein the text includes metadata corresponding to the range of the transcription.

28. The system of claim 21, wherein when retrieving a portion of the audio signal, the processor is configured to:

initiate a media player, and  
use the media player to obtain the portion of the audio signal.

29. The system of claim 28, wherein the media player is configured to:

identify the portion of the transcription, and  
retrieve the portion of the audio signal corresponding to the portion of the transcription.

30. The system of claim 29, wherein when identifying the portion, the media player is

configured to:

identify time codes associated with a beginning and an ending of the portion of the transcription.

31. The system of claim 29, wherein the portion of the transcription includes a starting position in the transcription; and

wherein when identifying the portion, the media player is configured to:

identify a time code associated with the starting position in the transcription.

32. The system of claim 21, wherein when providing the portion of the transcription and the portion of the audio signal to the user, the processor is configured to:

present a split screen in a translation window, the translation window including a translation section and a transcription section, and

display the portion of the transcription in the transcription section.

33. The system of claim 21, wherein when providing the portion of the transcription and the portion of the audio signal to the user, the processor is configured to:

visually synchronize the providing of the portion of the audio signal with the portion of the transcription.

34. The system of claim 33, wherein the portion of the transcription includes time

codes corresponding to when words in the transcription were spoken.

35. The system of claim 34, wherein when visually synchronizing the providing of the portion of the audio signal with the portion of the transcription, the processor is configured to:

compare times corresponding to the providing of the portion of the audio signal to the time codes from the portion of the transcription, and

visually distinguish words in the portion of the transcription when the words are spoken during the providing of the portion of the audio signal.

36. The system of claim 21, wherein when providing the portion of the transcription and the portion of the audio signal to the user, the processor is configured to:

permit the user to control the providing of the portion of the audio signal.

37. The system of claim 36, further comprising:

foot pedals configured to aid the user to at least one of fast forward, speed up, slow down, and back up the providing of the portion of the audio signal.

38. The system of claim 36, wherein when permitting the user to control the providing of the portion of the audio signal, the processor is configured to:

permit the user to rewind the portion of the audio signal at least one of a predetermined

amount of time and a predetermined number of words.

39. The system of claim 21, wherein the processor is further configured to:  
publish the translation to a user-determined location.

40. A graphical user interface, comprising:  
a transcription section that includes a transcription of non-text information in a first language;  
a translation section that receives a translation made by the user of the non-text information into a second language; and  
a play button that, when selected, causes:  
retrieval of the non-text information to be initiated,  
playing of the non-text information, and  
the playing of the non-text information to be visually synchronized with the transcription in the transcription section.

41. The graphical user interface of claim 40, wherein the transcription visually distinguishes names of people, places, and organizations.

42. The graphical user interface of claim 40, further comprising:  
a configuration button, that when selected, causes a window to be presented, the window

permitting an amount of backup to be specified, the amount of backup including one of a predetermined amount of time and a predetermined number of words.

43. The graphical user interface of claim 42, wherein the window further permits a name to be given for the translation and a location of publication to be specified.

44. The graphical user interface of claim 40, wherein the play button further causes words in the transcription to be visually distinguished in synchronism with the words in the non-text information being played.

45. The graphical user interface of claim 40, wherein the non-text information includes at least one of audio and video.

46. The graphical user interface of claim 40, wherein the graphical user interface is associated with a word processing application.

47. A method, comprising:  
a user listening to an audio playback of information in a first language while viewing a textual transcription of said information in said first language on a transcription section of a graphical user interface (GUI), said textual transcription being synchronized with said audio playback; and



said user translating said audio playback of said information thereby obtaining a translation in a second language, said user using a different section of said GUI to display said translation while making said translation,

whereby the synchronizing of said audio playback with said textual transcription aids said user in making said translation.

**IX. EVIDENCE APPENDIX**

none

**X. RELATED PROCEEDINGS APPENDIX**

None